DOCUMENT RESUME

ED 435 192 FL Q26 045

AUTHOR	Basci, Pelin
TITLE	Beyond Classroom Achievement: Standardized Turkish Tests.
ISSN	ISSN-1095-8096
PUB DATE	1999-00-00
NOTE	6p.
AVAILABLE FROM	AATT Bulletin, Portland State University, Dept. of Foreign Languages and Literatures, P.O. Box 751, Portland, OR 97207.
PUB TYPE	Journal Articles (080) -- Reports - Descriptive (141)
JOURNAL CIT	AATT Bulletin; n23-24 p14-17 Spr-Fall 1999
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Higher Education; *Language Tests; Professional Associations; Program Descriptions; Rating Scales; Scoring; *Standardized Tests; *Test Construction; Test Use; *Turkish; Uncommonly Taught Languages

ABSTRACT

The goals, concerns, and issues addressed by a committee of the American Association of Teachers of Turkic Languages (AATT) in the process of developing a standardized Turkish language proficiency test at the college level are examined. The testing committee aimed at incorporating recent scholarship on second language acquisition, teaching methods, and assessment into development of a test for intermediate and advanced proficiency levels. Scoring was designed to indicate both a competitive (rating) score and a proficiency level, to meet both diagnostic and placement needs. The test incorporates visual and authentic materials. Considerations in test administration and in test reliability and validity are discussed briefly. (MSE)

Reproductions supplied by EDRS are the best that can be made
from the original document.

BEYOND CLASSROOM ACHIEVEMENT: STANDARDIZED TURKISH TESTS

Pelin Başci
Portland State University

ED 435 192

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Pelin Başci
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

FL026045

BEYOND CLASSROOM ACHIEVEMENT: STANDARDIZED TURKISH TESTS

Pelin Başcı
Portland State University

An increasing separation of tasks has taken place in many "commonly taught" second or foreign languages such as English and Spanish, where the test builder is no longer necessarily the instructor. This is, of course, especially true for standardized tests, although the same principle also applies to oral proficiency assessment.[1] As a matter of fact, for some of these language areas, assessment is both an independent field of research and a full-fledged industry. There are advantages and disadvantages to such expansions in any field: the advantage stems from the professional work committed to assessment research and implementation, while the disadvantage most remarkably arises from the commercial interests built around highly standardized tests such as the Test of English as a Foreign Language (TOEFL). In some of these cases, special crash courses, study books, and exercise tapes all come to function as parts of a vast enterprise.

One doubts seriously that standardized tests in Turkish will ever become commercialized to the extent of, for example, TOEFL. While this may indeed be the good news, the bad news is that testing and assessment of Turkish within the US context has hardly reached the widespread professionalism that underlies the assessment research, test preparation, and implementation in some of the commonly taught language areas. No doubt, this is also due, at least in part, to the status of Turkish as a non-cognate critical language in a largely English-speaking world: there is no comparison between the number of students who take Turkish and those who take, for example, Spanish or other Romance languages in the US. Both the demand for standardized tests and the corpus of any data that result from such tests and would inform assessment research are smaller. Moreover, in many US academic institutions, Turkish instructors are asked to strike a difficult balance between their increasingly diversified academic and teaching responsibilities and limited resources.

One of the main goals of the testing committee for Turkish, which was founded in the late 1990s, has been to broaden the scope of interest and research in the testing and assessment of Turkish, and thereby promote professional awareness of the topic. It is with these broader goals in mind that I would like to expand further on the questions that the testing committee had to address in preparing model intermediate and advance tests for the 1998 program of the American Research Institute in Turkey (ARIT) for language study in Boğaziçi University.

Second language acquisition research, particularly the debate that centered around the distinction between learning and acquisition,[2] regardless of its shortcomings particularly for non-cognate languages,[3] provided useful analytic categories for rethinking classroom instruction. Debates about the place of comprehensible input produced in meaningful authentic contexts, the presence or absence of the "monitor" as a "grammar police" in learners' minds, and the delicate balance between accuracy and communicative competence have been informing foreign language methodology as well as assessment.[4] The preparation of multi-dimensional tests, rather than one-dimensional paper-and-pen examinations, incorporating authentic material that has some appeal and relevance to the test-taker's needs, and designing assessment models that are user-friendly and tests that help students rather than keeping them in their place, have been accompanying debates of instructional methodology.[5] Moreover, with the national move towards clearly defined proficiency goals as the organizing principle in teaching, "knowing" a second or foreign language

has been re-defined as the ability to perform, "to do" things, to carry out diverse tasks using all four language skills in the authentic environment of the target language and culture.

The AATT testing committee aimed at incorporating much of the useful recent scholarship on acquisition, methods, and assessment into the process of developing model tests for two different proficiency levels. In this process, the Turkish proficiency guidelines set the framework for developing the model tests. However, as the committee was charged with the task of creating models for the ARIT test, the question of intent had to be addressed: to what degree were models, like those prepared for ARIT, proficiency assessment tests? For the models to be successful, other significant variables had to be accounted for. Prepared within the framework of proficiency guidelines, the models had to become a point of departure for reliable standardized tests that best suited the needs of ARIT and reflected the educational context that produced the intermediate and advanced students of Turkish who would take these tests. For the testing committee, the identification and recognition of all the variables and different criteria in the construction of models were of great significance, for they helped to define what Bachman calls "the abilities we wish to measure and the means of measurement."[2]

Tests—while by no means the only criterion of evaluation—help ARIT decide what student to select for an in-country study program and whom to reward with a scholarship. In this sense, the models had to address simultaneously the "selection," "readiness," and "entrance" goals. Meanwhile, there was also a greater need for the tests to inform instructors and program administrators with increased accuracy about the students' readiness to pursue intermediate and advanced study in the program, which in turn, required that the tests "diagnose" student levels and reduce the burden of further "placement" procedures.

In addition to selection, readiness, and entrance goals, the tests' ability to diagnose and place students into expected levels of proficiency implied that the contexts, functions, and levels of accuracy to be tested should reflect the expectations of the receiving program. This brought to the fore further issues regarding instructional articulation, the implied continuity of levels, and expectations among different programs both within and without the US context—an issue that plays a part in decision-making, but is beyond the scope of any testing committee's immediate work.

Having listed some of the goals, major concerns, and issues to be addressed, it would be thoroughly unrealistic to claim with any level of certainty that all of these issues were resolved and that "perfect" models for standardized national tests were in fact created by the Turkish testing committee. Yet the following may shed light on how the committee handled some of these concerns.

(1) Addressing the objectives: The committee designed the scoring to indicate both a competitive score and a proficiency level in order to meet the different set of goals mentioned above. This meant that the tests would indicate both a demonstrated level of proficiency—as in intermediate-mid or intermediate-high—and a numerical score showing a ranking of the student in relation to other competitors for selection purposes.

In designing the questions, the committee made a serious effort to address different proficiency levels for different skills. For example, in the section that tested speaking, intermediate students were asked to communicate concrete, descriptive information about the self and their immediate environment; deliver this information through a range of speech conventions such as questions and commands, using grammatical structures such as "var/yok", and the future, present continuous, past, and aorist tenses; and display a certain amount of control and appropriateness in their use

of these basic structures. Their functional skills were put to test in well-defined tasks that related to the expected benchmarks in intermediate proficiency levels, such as inviting friends over for dinner and providing them with directions and ordering at a restaurant. Advanced students were asked to display the ability to talk about the self in a more detailed and specialized fashion, incorporating their career options, future plans, and a discussion of cultural differences. The situational contexts in which their functional skills were tested include, but are not limited to, retrieving stolen valuables such as a passport, making grievances at a hotel while employing appropriate structures and increased awareness of the socio-linguistic rules of Turkish.

(2) Constructing a multi-dimensional test while addressing issues of scoring: In addition to a written text, the tests included a variety of visuals in the form of pictures, photographs, newspaper clippings, and an audio section which contained questions read on a tape and required answers spoken into a tape. The model tests were constructed to measure student skills at advanced and intermediate levels, separately examining all four areas—speaking, listening, reading, and writing—even though the committee felt strongly that each of these seemingly separate receptive and productive skills contributed to one another. At least in the case of speaking and listening comprehension the committee therefore designed the scoring guides accordingly. A certain percentage of the scores received in listening contributed, for example, to the student's scores in speaking and vice versa.[3]

(3) The use of authentic material: In the testing of different skills, including reading and writing, the committee tried to provide students with communicative contexts such as filling in a subscription form (intermediate) and processing information from a film review (advanced). In all of the questions in each sub-section, the committee relied to a large extent on realia: photos and cartoons from newspapers rather than drawings, newspaper clippings rather than composed or altered written texts, and unscripted conversations depicting a certain negotiation of meaning such as a rendezvous, to name a few.[4]

(4) Reliability and validity: In order to increase the reliability of the model tests, attempts were made, even if limited in scope, to administer "pre-testing" on those groups who could not themselves participate in the "competition." Often, time constraints more than anything else hampered further work in this direction. Moreover, despite all that is said and done, any assessment process itself can only be assessed when there is additional input on implementation, which is to say the actual administration process.

Finally, it is the input coming from the implementation process that will help to improve the quality and reliability of such models for future use. Were the instructions to the students and to the proctor clear enough? How did the students perform with the time restrictions, since the tests were not constructed as "speed-tests"? What role did the physical conditions and limits of the testing environment play? Even the identity of the proctor could play a significant role in the implementation of any test. In this case, a Turkish-speaking instructor who would have an idea of the test material, realize what is being tested, and could handle unexpected situations regarding the test, is certainly preferable to a substitute with no functional language abilities in Turkish and no training in any aspect of teaching and testing. Similarly, the quality and availability of tape recorders at a given institution, the conditions of the room in which the test was administered—all of these and other variables could impact the reliability of any standardized test. It is particularly, but not exclusively, on this note about implementation that the testing committee would like to solicit responses to the model tests and invite a larger participation

in the debate on the assessment of Turkish as a foreign and/or second language.

NOTES

[1] There are those who believe that oral-proficiency assessment, particularly in the form of an interview, works best if the interviewer is not the teacher.

[2] For a sample of Krashen's prolific work, see Stephen Krashen, *Second Language Acquisition and Second Language Learning* (New York: Pergamon Press, 1981).

[3] For a critique of Krashen and the monitor model, see Ronald M. Barasch and C. Vaughan James, ed., *Beyond the Monitor Model: Comments on Current Theory and Practice in Second Language Acquisition* (Boston, Massachusetts: Heinle & Heinle Publishers, 1994).

[4] Theodore V. Higgs, ed., *Teaching for Proficiency, the Organizing Principle* (Lincolnwood: National Textbook Company, 1989).

[5] Andrew D. Cohen, *Assessing Language Ability in the Classroom* (Boston, Massachusetts: Heinle & Heinle Publishers: 1994).

[5] Lyle Bachman, *Fundamental Considerations in Language Testing* (New York: Oxford University Press, 1990), 81.

[6] For the scoring guide please, see Güliz Kuruoğlu, "Turkish Proficiency Tests: A National Model" in this issue of the *Bulletin*.

[7] This would be one example in which certain justifiable compromises were made in the use of the material. To maintain high-quality audibility, recordings were produced at a later time in a studio environment, although every effort was made during this process to maintain the natural flow of the original unscripted conversation.